The Behavior Analyst 1999, 22, 149–157 No. 2 (Fall)

In Response

Quantitative Integration of Single-Subject Studies: Methods and Misinterpretations

Scott H. Kollins
Western Michigan University

M. Christopher Newland
Auburn University

Thomas S. Critchfield
Illinois State University

Derenne and Baron (1999) criticized a quantitative literature review by Kollins, Newland, and Critchfield (1997) and raised several important issues with respect to the integration of single-subject data. In their criticism they argued that the quantitative integration of data across experiments conducted by Kollins et al. is a meta-analysis and, as such, is inappropriate. We reply that Kollins et al. offered behavior analysts a technique for integrating quantitative information in a way that draws from the strengths of behavior analysis. Although the quantitative technique is true to the original spirit of meta-analysis, it bears little resemblance to meta-analyses as currently conducted or defined and offers behavior analysts a potentially useful tool for comparing data from multiple sources. We also argue that other criticisms raised by Derenne and Baron were inaccurate or irrelevant to the original article. Our response highlights two main points: (a) There are meaningful quantitative techniques for examining single-subject data across studies without compromising the integrity of behavior analysis; and (b) the healthiest way to refute or question findings in any viable field of scientific inquiry is through empirical investigation.

Psychologists, who agree on little, widely endorse the proposition that, in the study of behavior, methods matter. This resonant concordance brings unity to an otherwise chaotic world of contrasting scholarly perspectives. The mantra "methods matter" should be repeated often, but especially so in developing disciplines, in which procedures have yet to become standardized and research questions remain to be clarified.

All authors contributed equally to this work. Order of authorship on this paper was thus determined on the basis of order from the previously published paper. Work on this manuscript was supported in part by a Faculty Research Development Award from Western Michigan University (S.H.K.) and Grant ES 06466 from the National Institute of Environmental Health Sciences (M.C.N.).

Address correspondence to Scott H. Kollins, Department of Psychology, Western Michigan University, Kalamazoo, Michigan 49008-5052 (E-mail: scott.kollins@wmich.edu).

The experimental analysis of human behavior is one such developing field. Human studies are published and cited infrequently in the experimental analysis of behavior (Dougherty, 1994; Hyten & Reilly, 1992; Perone, 1985), and the range of topics addressed by these studies is limited (Doughterty, 1994). Among primary empirical studies published in Volumes 63 through 70 (1995 to 1998) of the *Journal of the Experimental Analysis of Behavior* (*JEAB*), only about 17% (27 of 162) involved human subjects. Compounding the problems arising from a limited data base is the discomfiting presence of substantial unexplained variability in findings from the experimental analysis of human behavior, including atypical subjects within studies, failures to replicate findings from animal studies, and differential outcomes across laboratories. These problems and their implications have been addressed extensively (e.g., Baron & Perone, 1982; Baron, Perone, & Galizio, 1991a, 1991b) and need not be repeated here. For present purposes, the most salient and troubling consequence of this state of affairs is that some have come to regard unexplained variability as one of the hallmarks of human research (Branch, 1991; Shull & Lawrence, 1991) and perhaps of human subjects in general. Such conclusions are damaging to the still-fledgling movement because, left unrebutted, they portray the experimental analysis of human behavior as an enterprise with limited potential to advance a general science of behavior. Perhaps worse, such conclusions erode the general credibility of behavior analysis, which has placed great stock in extending principles derived from nonhuman species to the interpretation and alteration of human behavior.

Against this backdrop, Kollins, Newland, and Critchfield (1997) undertook a review of studies relevant to human choice as described by Baum's (1974) generalized matching law. The review revealed and quantified substantial variability in the degree to which relevant quantitative models have fit human behavior, and in estimates of human sensitivity to reinforcement derived from the models. Its general conclusions agreed with a qualitative review conducted 14 years previously (Pierce & Epling, 1983). The primary innovation of the Kollins et al. review lay in its introduction of a quantitative approach to estimating the contributions of procedural factors to the variability in choice outcomes. Kollins et al. concluded that the existing evidence points to experimenters, not human subjects, as the source of variability in the human choice literature. That is, methods matter, and the findings suggest some intriguing hypotheses about the way that experimental procedures influence human choice.

Recently, Derenne and Baron (1999) criticized the Kollins et al. (1997) review at some length, objecting to the review on two grounds. First, they described the quantitative approach used by Kollins et al. as meta-analysis, a technique that is conventionally linked to Fisherian comparisons across groups and thus is particularly unsuitable for use with single-subject data. Second, they dismissed certain findings of the review as improbable, attributing them to use of flawed techniques for comparing across studies. We address these claims below.

Quantitative Comparisons Across Studies

The Kollins et al. (1997) analyses were rooted in Baum's (1974) generalized matching equation, which describes the positive linear relationship of relative response rate to relative reinforcement rate when plotted on logarithmic coordinates. The slope parameter of Baum's equation is conventionally regarded as an estimate of sensitivity to reinforcement (Baum, 1974; Davison & McCarthy, 1988). Emphasizing this parameter, we employed a four-step process. First, 310 slope values from individual subjects and one slope value derived from an average of several subjects were obtained from 25 published studies.[1] Second, a frequency distribution of slopes was constructed and compared with one representing nonhumans, as summarized by Baum (1979). Overall, human slopes tended to be shallower than those of nonhumans (median = 0.70 for humans, 0.85 for nonhumans), and the distribution was more variable (interquartile range = 0.31 to 1.01 for humans, 0.75 to 0.97 for nonhumans).

Third, distributions of slopes were examined quantitatively and graphical-

[1] Slopes typically were described in studies employing concurrent-schedule procedures. Studies in which a single response rate is reported can be described by Herrnstein's (1970) single-schedule hyperbolic equation, which includes an "other reinforcement" free parameter that permits an estimate of sensitivity to reinforcement in terms of the generalized matching law. See Kollins et al. (1997) for details.

TABLE 1

Slope values from human studies reviewed by Kollins et al. (1997) and from nonhuman studies reviewed by Baum (1979).

Percentile	Human slope value	Nonhuman slope value
10th	0.07	0.66
20th	0.25	0.73
30th	0.45	0.77
40th	0.58	0.82
50th (median)	0.70	0.85
60th	0.87	0.90
70th	0.94	0.94
80th	1.12	1.01
90th	1.26	1.07

ly, applying a technique called the *empirical quantile-quantile* (Q-Q) plot[2] (Cleveland, 1985; Newland, 1997) to compare the distributions of data from studies that differed along certain methodological dimensions. Kollins et al. (1997) applied this approach to aggregate data across experiments while preserving the integrity and power of behavior analysis in its ability to account for the behavior of individual subjects. Because of the novelty of the application, we take the liberty of illustrating it here.

In the most global comparison, nine deciles (10th, 20th, . . . , 90th percentiles) representing a distribution of slopes taken from animal studies (from Baum, 1979) were plotted against nine deciles describing the distribution of slopes from the human studies that we reviewed. Table 1 lists the raw deciles, and Figure 1 graphically depicts their relationship. In this representation it is clear that the deciles of the different distributions fall along a straight line that can be described by $Y = a + mX$. If $a = 0$ and $m = 1$, then the distributions are identical; to the extent that $a \neq 0$ and $m \neq 1$, the distributions differ from one another. This formed the basis for the statistical comparisons made by Kollins et al. (1997): Was a different from 0 and m different from 1? The least squares regression line in Figure 1 (solid line) is described by $Y = 2.94X - 1.83$, and it is quite different from the major diagonal (dotted line) represented by $Y = 1*X + 0$ (for details, see Kollins et al.).

In this method of display, a nonunity slope indicates that the two distributions differ from one another and provides information about the dimensions along which they differ, only two of which will be highlighted here (see Cleveland, 1985, or Newland, 1997, for further detail). First, the magnitude of difference between the distributions depends on what portion of the distribution is being considered, an interpretation similar to a statistical interaction. Second, the two distributions cover different ranges of values. In this case, the distribution represented on the vertical axis is more variable (the extremes are farther apart) than that represented on the horizontal axis. This is necessarily the case if the slope is greater than 1.0 (a slope less than 1.0 would indicate that the distribution on the horizontal axis is the more variable one). Additional information is provided by the intercept, the magnitude of the slope, and the linearity of the line, among others (see Cleveland, 1985, and Newland, 1997, for additional details).

The fourth and most important step taken by Kollins et al. (1997) was to employ Q-Q plots to forge comparisons among human studies employing different types of procedures. Kollins et al. found that the overall variability in matching slopes was related to a variety of methodological factors, such as the use of monetary versus other consequences, measurement of button press versus other responses, and the presence or absence of schedule-correlated stimuli. That is, methods matter, and the results of Kollins et al. advanced testable hypotheses about the ways in which experimental procedures influence human choice.

[2] Sometimes this is called a *percentile-percentile* plot. The difference is that quantiles range from 0 to 1 and percentiles range from 0 to 100.

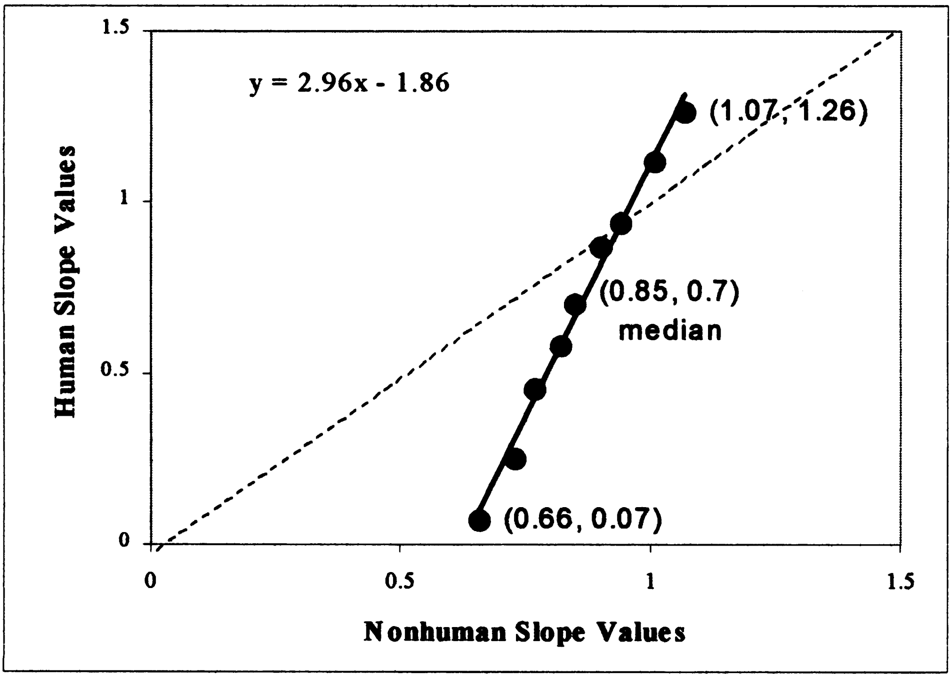

Figure 1. Q-Q plot of distribution depicted in Table 1. Each decile of the human slope distribution (as reported by Kollins et al., 1997) is plotted as a function of the same decile of the nonhuman slope distribution (as reported by Baum, 1979). The regression equation provides information regarding the slope and intercept and can be interpreted as a measure of how the distributions differ from unity ($Y = 1*X + 0$). For clarity, the values for the 10th, 50th (median), and 90th deciles are labeled. Redrawn based on Figure 3 from Kollins et al.

Was Kollins et al. (1997) a Meta-Analysis?

Like methods, terms also matter. The issue of whether the Kollins et al. (1997) paper was a meta-analysis is an important one for the same reasons that any application of a technical term is important. Functionally, the answer depends on how the term is defined and used in the contemporary literature. After calling Kollins et al. a meta-analysis, Derenne and Baron (1999) provided a synopsis of standard criticisms commonly applied to such reviews. If the review was indeed a meta-analysis, then these concerns apply; if it was not a meta-analysis, then their criticism bears a second look.

Originally, Glass, McGaw, and Smith (1981) introduced the term *meta-analysis* to describe a *general quantitative strategy* for seeking patterns in a research literature. They defined a meta-analysis as follows:

> By recording the properties of studies and their findings in quantitative terms, the meta-analysis of research invites one who would integrate numerous and diverse findings to apply the full power of statistical methods to the task. Thus, it is not a technique; rather it is a perspective that uses many techniques of measurement and statistical analyses. (p. 21)

It was in this sense that Kollins et al. (1997) described the review as "a form of meta-analysis" (p. 212). In this sense, the appellation applies as well to acknowledged classics in the literature describing the matching law, including Baum (1979) and Myers and Myers (1977).

Glass et al.'s (1981) intent notwithstanding, as ordinarily used the term *meta-analysis* now describes a specific technique that bears little resemblance to the approach used by Kollins et al.

(1997). The review captured the original intent of meta-analyses as described in the quotation above, but had it been called a meta-analysis, the claim would have been criticized. Current discussions of the term entail attempts to narrow its definition so that there is consistency in how the term is used (e.g., Hunter & Schmidt, 1990; Wachter & Straf, 1990). In these sources, the term applies to formal efforts that (a) synthesize research using (b) a quantitative description of effect that (c) includes some measure of central tendency (almost always a mean, a proportion, or an odds ratio). These are (d) expressed in the context of a measure of variability (almost always a standard deviation or confidence interval) in which (e) the unit of analysis is the group or the research report. Note that, together, items c and d are used to calculate the "effect size," which lies at the conceptual heart of modern meta-analytic techniques (Glass et al., 1981; Hunter & Schmidt, 1990).

The Kollins et al. (1997) review satisfies the first two criteria—data were synthesized across studies and a quantitative description of the relations among such data was used. We agree with Derenne and Baron (1999) that thorny issues arise when findings of different studies are considered together. But we reserve judgment on Derenne and Baron's assertion that data aggregation is especially troublesome for single-subject research. This matter has been debated previously at great length (Allison & Gorman, 1993; Busk & Serlin, 1992; Gingerich, 1984; Salzburg, Strain, & Baer, 1987; Scruggs & Mastropieri, 1998; Scruggs, Mastropieri, & Casto, 1987a, 1987b, 1987c; White, 1987; White, Rusch, Kazdin, & Hartmann, 1989). For the present, we note simply that, although methods certainly matter for literature reviews as they do for primary empirical studies, standardized methods do not exist for integrative literature reviews as they do for experiments (Mosteller, 1990). Until such time as they do, no literature review is ever beyond reproach on general grounds, because *all* efforts to combine data across experiments are compromised by necessarily subjective decisions about procedures for study inclusion, data aggregation, and so forth.

Particularly with respect to the third, fourth, and fifth criteria, the Kollins et al. (1997) review accomplished goals that meta-analyses, as conventionally understood, cannot. This is because the technique in question, Q-Q plots, uses as its unit of analysis distributions of individual data, rather than effect sizes that are calculated from aggregated measures of central tendency and variability (that are necessarily derived from combining data across subjects and raise conceptual problems for the analysis of behavior; e.g., Sidman, 1960). This analytical strategy is reflected in the graphical display of the data as well as in the regression analyses. The slope is a central feature of the Q-Q plot analysis because it incorporates a comparison of two entire *distributions*—distributions that comprise individual-subject data.

We make no "best practice" claims about Kollins et al. (1997). To date, behavior analysts have not undertaken quantitative syntheses of experimental literatures frequently enough to shed light on what constitutes best practice. We suggest only that the approach taken by Kollins et al. is consistent with fundamental tenets of behavior analysis because it maintains information at the level of the individual subject; it emphasizes rather than obscures variability; and it focuses on potential determinants of behavior as defined in the experimental analysis of behavior. In summary, the Kollins et al. (1997) review offers behavior analysts one alternative to the modern practice of meta-analysis.

On Unexpected Findings

With regard to our specific findings, Derenne and Baron (1999) focused on those that they deemed "puzzling" (p. 39), "surprising" (p. 35), "unexpected" (p. 35), and "remarkable" (p. 36). For example, studies conducted with nonhumans indicate that sensitivity to reinforcement in choice is enhanced by schedule-correlated discriminative stimuli; the Kollins et al. (1997) review suggested otherwise for humans. Clearly, unusual results deserve unusual scrutiny, but Derenne and Baron at times mischaracterized how emphatically the results were presented and, in the process, adopted a surprisingly narrow perspective on the possibilities raised by unexpected findings.

In the former case, Derenne and Baron (1999) asserted that findings were presented unequivocally, without reference to the limitations of our methods. Yet the review's major focus was on the importance of methods, as illustrated by the concluding comment: "We should strive for a more thorough understanding of how the procedures used to study behavior—of all species, but especially of humans—affect the behavior under question" (Kollins et al., 1997, p. 218). In addition, it was carefully noted that

> Because most studies differ procedurally in several ways, the binary comparisons conducted here are artificially simple in that they do not reflect the possible interaction among factors that could influence sensitivity to reinforcement. . . . Many other kinds of comparisons are possible, as the relevant studies differ procedurally on many dimensions. (Kollins et al., p. 215)

Yet Derenne and Baron worried that "Because qualifications were *not* expressed in the article, the implication is that each of the variables that influenced human sensitivity [to reinforcement] operated independently" (p. 38; italics added). We encourage readers of the comment to consult our original work and draw their own conclusions about how the results were presented.

With respect to results, Derenne and Baron (1999) argued that these unexpected findings must be artifacts of data manipulation, because they "run counter to both theory and common observation" (p. 36). Yet, the absence of citations makes it impossible to determine which theory and observation Derenne and Baron considered most relevant to these unexpected findings. On the basis of previous position papers by Baron and colleagues (e.g., Baron & Perone, 1982; Baron et al., 1991a, 1991b), we infer that Derenne and Baron share our bias for assuming interspecies continuity of behavior principles, and our respect for the prodigious data base accumulated through past behavioral research. We note, however, that, given a relative paucity of human operant research (e.g., Hyten & Reilly, 1992), the literature relevant to most operant research questions tends to consist largely of studies conducted with nonhumans. And unqualified confidence in this source of guidance has its pitfalls.

For example, referring to one of the findings reported by Kollins et al. (1997), Derenne and Baron (1999) stated that "it seems remarkable . . . that sensitivity [to reinforcement] should be reduced by adding discriminative stimuli" (p. 36). This finding does indeed seem at odds with findings from the nonhuman laboratory (e.g., Davison & McCarthy, 1988). Two of the human studies reviewed, however, explicitly manipulated the manner in which schedule-correlated discriminative stimuli were presented (Horne & Lowe, 1993; Takahashi & Iwamoto, 1986). In both of these studies, the presence of discriminative stimuli alone did not categorically result in behavior that was better described as matching (i.e., more sensitive to the changing contingencies), despite what might be expected based on "theory and common observation" from the nonhuman literature.

The preceding example serves as a reminder that two conclusions are possible when human results deviate from those obtained from nonhumans: (a) Impressions of human behavior (which, if the truth were known, functions like that of other species) have been corrupted in some way by idiosyncrasies of the research methods, or (b) human behavior is, in fact, different from that of other species. The latter possibility has been the subject of much controversy within behavior analysis. Fortunately, science offers tried and true means for resolving conflict, and, in the present case, far more important than knowing the correct answers about human choice is a commitment to seeking those answers in a way that promotes the long-term health of a science of human behavior.

Over the years, no one has argued more forcefully than Baron and his colleagues that points of divergence between human and animal data should be viewed as a research mandate (Baron & Perone, 1982; Baron et al., 1991a, 1991b). In this regard, we find it ironic that the Derenne and Baron (1999) commentary did not provide new empirical findings that inform the field about human choice. But solutions to the current impasse are self-evident. Derenne and Baron suggested that improved methods of literature review might lead to findings more in accord with the animal literature. Such a reanalysis of the data would be a welcome contribution. In addition, Kollins et al. (1997) identified many testable hypotheses about the role of procedural variations on human choice. The relevant experiments would not be difficult to conduct.

Historical examples suggest that, although the answers regarding "aberrant" human behavior are not simple or easily obtained, efforts to seek them empirically will prove to be valuable. Consider the well-known case of human fixed-interval (FI) performance, which rarely reflects the positively accelerating "scalloped" pattern of responding so widely attributed to nonhumans performing on similar schedules (e.g., Weiner, 1969, 1972). This outcome has spawned diverse hypotheses. Perhaps incidental factors (e.g., prior learning) confound human FI performance. Perhaps uniquely human characteristics (e.g., language) are operating. Perhaps the overall pattern (scalloping) is an artifact of more fundamental processes (Zeiler, 1984) to which both human and nonhuman laboratories should attend. These hypotheses have stimulated insights into important topics such as instructional control (Baron & Galizio, 1983), behavioral history (Wanchisen, Tatham, & Mooney, 1989), and the alleged ubiquity of FI scalloping in nonhumans (Hyten & Madden, 1993). But empirical work (both experimentation and the reanalysis of existing data) was the primary foundation of this progress.

In summary, an empirical approach to problem solving is the bedrock of an emerging area like the experimental analysis of human behavior, in which both procedures and findings are subject to competing interpretations. Sadly, in a field that is desperately short of workers (e.g., only five laboratories accounted for nearly half of the human studies published in *JEAB* during 1995–1998), vigorous empirical efforts can be difficult to muster. Whatever its possible shortcomings, the Kollins et al. (1997) review succeeded in organizing a diverse human choice literature, identifying heretofore unrecognized patterns in the data, and generating compelling research questions based upon the data. Although ever faithful to the maxim that methods matter, Derenne and Baron (1999), in their haste to dismiss outcomes that troubled them, may have temporarily lost sight of a second research maxim: Data speak louder than words. Will each of the findings of Kollins et al. hold up to further empirical scrutiny? That seems unlikely. But which findings may be spurious and which may provide genuine insight into human behavior is a matter to be decided in a court in which data constitute the jury.

REFERENCES

Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behaviour Research, & Therapy, 31,* 621–631.

Baron, A., & Galizio, M. (1983). Instructional control of human operant behavior. *The Psychological Record, 33,* 495–520.

Baron, A., & Perone, M. (1982). The place of the human subject in the operant laboratory. *The Behavior Analyst, 5,* 143–158.

Baron, A., Perone, M., & Galizio, M. (1991a). Analyzing the reinforcement process at the human level: Can application and behavioristic interpretation replace laboratory research? *The Behavior Analyst, 14,* 95–105.

Baron, A., Perone, M., & Galizio, M. (1991b). The experimental analysis of human behavior: Indispensable, ancillary, or irrelevant? *The Behavior Analyst, 14,* 145–155.

Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior, 22,* 231–242.

Baum, W. M. (1979). Matching, undermatching, and overmatching in studies of choice. *Journal of the Experimental Analysis of Behavior, 32,* 269–281.

Branch, M. N. (1991). On the difficulty of studying "basic" behavioral processes in humans. *The Behavior Analyst, 14,* 107–110.

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single case research design and analysis: New directions for psychology and education* (pp. 187–212). Hillsdale, NJ: Erlbaum.

Cleveland, W. S. (1985). *The elements of graphing data.* Pacific Grove, CA: Wadsworth and Brooks/Cole.

Davison, M., & McCarthy, D. (1988). *The matching law.* Hillsdale, NJ: Erlbaum.

Derenne, A., & Baron, A. (1999). Human sensitivity to reinforcement: A comment on Kollins, Newland, and Critchfield's (1997) quantitative literature review. *The Behavior Analyst, 22,* 35–42.

Dougherty, D. M. (1994). The selective renaissance of the experimental analysis of human behavior. *The Behavior Analyst, 17,* 169–174.

Gingerich, W. J. (1984). Meta-analysis of applied time-series data. *Journal of Applied Behavioral Science, 20,* 71–79.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior, 13,* 243–266.

Horne, P. J., & Lowe, C. F. (1993). Determinants of human performance on concurrent schedules. *Journal of the Experimental Analysis of Behavior, 59,* 29–60.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

Hyten, C., & Madden, G. J. (1993). The scallop in human fixed-interval research: A review of problems with data description. *The Psychological Record, 43,* 471–500.

Hyten, C., & Reilly, M. P. (1992). The renaissance of the experimental analysis of human behavior. *The Behavior Analyst, 15,* 109–114.

Kollins, S. H., Newland, M. C. & Critchfield, T. S. (1997). Human sensitivity to reinforcement in operant choice: How much do consequences matter? *Psychonomic Bulletin and Review, 4,* 208–220. Erratum: *Psychomonic Bulletin and Review, 4,* 431.

Mosteller, F. (1990). The future of meta-analysis. In K. W. Wachter & M. L. Straf (Eds.), *The future of meta-analysis* (pp. 185–190). New York. Russell Sage Foundation.

Myers, D. L., & Myers, L. E. (1977). Undermatching: A reappraisal of performance on concurrent variable-interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior, 27,* 203–214.

Newland, M. C. (1997). Quantifying the molecular structure of behavior: Separate effects of caffeine, cocaine, and adenosine agonists on interresponse times and lever press durations. *Behavioural Pharmacology, 8,* 1–16.

Perone, M. (1985). On the impact of human operant research: Asymmetrical patterns of cross-citation between human and nonhuman research. *The Behavior Analyst, 8,* 185–189.

Pierce, W. D., & Epling, W. F. (1983). Choice, matching, and human behavior: A review of the literature. *The Behavior Analyst, 6,* 57–76.

Salzburg, C. L., Strain, P. S., & Baer, D. M. (1987). Meta-analysis for single subject research: When does it clarify? When does it obscure? *Remedial and Special Education, 8,* 43–48.

Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior Modification, 22,* 221–242.

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987a). The quantitative synthesis of single subject research: Methodology and validation. *Remedial and Special Education, 8,* 24–33.

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987b). "The quantitative synthesis of single subject research: Methodology and validation": Reply to Owen White. *Remedial and Special Education, 8,* 40–42.

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987c). "The quantitative synthesis of single subject research: Methodology and validation": Response to Salzberg, Strain & Baer. *Remedial and Special Education, 8,* 49–52.

Shull, R. L., & Lawrence, P. S. (1991). Preparations and principles. *The Behavior Analyst, 14,* 133–138.

Sidman, M. (1960). *Tactics of scientific research.* New York: Free Press.

Takahashi, M., & Iwamoto, T. (1986). Human concurrent performances: The effects of experience, instructions, and schedule-correlated stimuli. *Journal of the Experimental Analysis of Behavior, 45,* 257–267.

Wachter, K. W., & Straf, M. L. (1990). *The future of meta-analysis.* New York. Russell Sage Foundation.

Wanchisen, B. A., Tatham, T. A., & Mooney, S. E. (1989). Variable-ratio conditioning history produces high- and low-rate fixed-interval performance in rats. *Journal of the Experimental Analysis of Behavior, 52,* 167–179.

Weiner, H. (1969). Controlling human fixed-interval performance. *Journal of the Experimental Analysis of Behavior, 12,* 349–373.

Weiner, H. (1972). Controlling human fixed-interval performance with fixed-ratio responding or differential reinforcement of low-rate responding in mixed schedules. *Psychonomic Science, 26,* 191–192.

White, D. M. (1987). "The quantitative synthesis of single subject research: Methodology and validation" comment. *Remedial and Special Education, 8,* 34–39.

White, D. M., Rusch, F. R., Kazdin, A. E., & Hartmann, D. P. (1989). Applications of meta-analysis in individual subject research. *Behavioral Assessment, 11,* 281–296.

Zeiler, M. D. (1984). The sleeping giant: Reinforcement schedules. *Journal of the Experimental Analysis of Behavior, 42,* 485–493.